

Mock LISA Data Challenges

Alberto Vecchio

On behalf of the LIST Task Force

<http://www.tapir.caltech.edu/dokuwiki/listwg1b:home>

Sixth LISA Symposium

Goddard, 19th - 23rd June 2006





Outline

- *Main goal of the talk: to announce the first round of Mock LISA Data Challenges (MLDCs) and provide the necessary information to participate*
- Some background information on MLDCs
- Details on MLDCs
- Next talk: MLDC tutorial



Some background

From e-mail by Bernard Schutz and Neil Cornish, co-chairs of LIST Working Group 1B (Data Analysis):

“At the December 2005 LISA International Science Team meeting in Pasadena the Data Analysis Working Group was more formally constituted, with co-chairs Neil Cornish and Bernard Schutz. The working group has, among other things, the task of organizing mock data challenges for the LISA data analysis development efforts.

The mock data challenges are the principal means we will have during the next couple of years to validate proposed algorithms and ensure that they can indeed meet the requirements of the LISA data analysis system. We envision that there will be a number of such challenges, starting with relatively simple ones and culminating (hopefully by the end of 2007) in a fairly realistic one.

To this end we are setting up a Mock Data Challenge Task Group”



Task Force members

Keith Arnaud	member
Stanislav Babak	member
John Baker	organizing secretary
Matthew Benacquista	member
Joan Centrella	member
Neil J. Cornish	WG1B co-chair
Curt Cutler	member
Shane L. Larson	member
B S Sathyaprakash	member
Michele Vallisneri	task force lead
Alberto Vecchio	task force lead
Jean-Yves Vinet	member



MLDC Task Force

- Charter:
 - Decide on a minimum set of physical conventions
 - Identify a minimum set of file descriptors
 - Decide on a data format
 - Develop a plan for conducting a series of progressively more difficult mock data challenges
 - Distribute training data sets
 - Decide on the metrics that will be used to measure the relative performance of different algorithms
 - Administer the mock data challenges
- Activities
 - Task force started to work in late February 2006
 - Fortnightly teleconferences (Mondays at 16:15 GMT, 8:15 PST) - everybody can join
 - Minutes, documents etc. are available at <http://www.tapir.caltech.edu/dokuwiki/listwg1b:home>
- **Today we are announcing Challenge-1: please join in!**



Goals of MLDCs

- Open issues in LISA data analysis:
 - Source specific
 - Global analysis
- Goals of MLDC:
 - To foster the **development of LISA data analysis**
 - To **demonstrate the technical readiness** already achieved by the gravitational-wave community in distilling a rich science payoff from the LISA data output
- Challenges are **non-competitive**, yet **blind tests**
- Each round of challenges consists of multiple data sets
- For each round we will distribute:
 - Challenge data sets (blind analysis - unknown parameters)
 - Training data sets (known parameters)



The challenge

Given a (mock) data set:

- I. Return the maximum amount of correct information about the sources that generate the GW signals present in the (mock) data set



The challenge

Given a (mock) data set:

- I. Return the maximum amount of correct information about the sources that generate the GW signals present in the (mock) data set
- II. Produce a technical note detailing the results, analysis method(s) and implementation



Challenge progression

- Challenge 1 (June 06 -> December 06)
 - Development and validation of building blocks of and basic tools for LISA data analysis
 - Analysis of data sets containing a single source or multiple non-overlapping sources (there is one exception though) in Gaussian and stationary instrumental noise and no foregrounds
- Challenge 2 (December 06 -> June 07)
 - Global analysis of data sets with minimum science in (more) realistic conditions. Data set will contain (details TBD):
 - Verification binaries
 - ~ 100 resolvable unknown binaries
 - A small band with strongly overlapping sources
 - A few MBH binaries (one of which at high SNR)
 - Galactic foreground



Challenge progression (cont')

- Challenge 2 cont' (December 06 -> June 07)
 - Tackle single source data sets containing
 - Extreme mass ratio inspirals
 - Short lived bursts
 - Stochastic signals
- Challenge 3 (June 07 -> December 07)
 - More realistic data sets
 - Larger number of sources and/or more general waveforms
 - More realistic noise (gaps, non-stationarity/Gaussianity)



C1 - Single source: WD binaries

- C1.1.1a
 - Source: **1 WD binary**
 - Waveform: monochromatic in source reference frame (no fdot)
 - Signal: $f \sim 1 \text{ mHz}$, SNR $\sim 10\text{-}20$
 - Noise: instrumental (Gaussian and stationary)
- C1.1.1b
 - Source: **1 WD binary**
 - Waveform: monochromatic in source reference frame (no fdot)
 - Signal: $f \sim 3 \text{ mHz}$, SNR $\sim 10\text{-}20$
 - Noise: instrumental (Gaussian and stationary)
 - Length data set: 1 yr
- C1.1.1c
 - Source: **1 WD binary**
 - Waveform: monochromatic in source reference frame (no fdot)
 - Signal: $f \sim 10 \text{ mHz}$, SNR $\sim 10\text{-}20$
 - Noise: instrumental (Gaussian and stationary)
 - Length data set: 1 yr



C1 - Multiple sources: WD binaries

- C1.1.2
 - Source: **5 (“real”) + 15 (mock) verification binaries**
 - Waveform: monochromatic in source reference frame (no fdot)
 - Signal: $f \sim 0.5 - 12$ mHz, $\text{SNR} > 10$
 - Noise: instrumental (Gaussian and stationary)
 - Length of data set: 1 yr
- C1.1.3
 - Source: **20 (known number) WD binaries**
 - Waveform: monochromatic in source reference frame (no fdot)
 - Signal: f over the whole band, $\text{SNR} > 10$.
 - Noise: instrumental (Gaussian and stationary)
 - Length of data set: 1 yr
- C1.1.4
 - Source: **~50 (unknown number) WD binaries**
 - Waveform: monochromatic in source reference frame (no fdot)
 - Signal: $f \sim 3$ mHz with $\Delta f \sim 15$ microHz, $\text{SNR} > 5$.
 - Noise: instrumental (Gaussian and stationary)
 - Length of data set: 1 yr

<http://astrogravs.nasa.gov/docs/ml>



C1 - Single source: MBH binary

- C1.2.1
 - Source: **1 MBH binary**
 - Waveform: restricted 2PN **in-spiral**
 - Signal: $\tau \sim 6$ months; $m_1 = U[1-5] \times 10^6 M_{\text{sun}}$ and $m_1/m_2 = U[1-4]$
SNR ~ 500
 - Noise: instrumental (Gaussian and stationary)
 - Length of data set: 1 yr
- C1.2.2
 - Source: **1 MBH binary**
 - Waveform: restricted 2PN **in-spiral**
 - Signal: $\tau \sim 14$ months; SNR ~ 20
 - Noise: instrumental (Gaussian and stationary)
 - Length of data set: 1 yr



C1/2 - Single source: EMRI

- C1.3.1
 - Source: **1 EMRI**
 - Waveform: kludge a la Barack & Cutler (2004)
 - Signal: $e = 0.2$ for $f_m = 2 \sim 3$ mHz, **SNR ~ 100**
 - Noise: instrumental (Gaussian and stationary)
 - Length of data set: **2 yr**
- C1.3.2
 - Source: **1 EMRI**
 - Waveform: kludge a la Barack & Cutler (2004)
 - Signal: 2 yr, $e = 0.2$ for $f_m = 2 \sim 3$ mHz, **SNR ~ 30**
 - Noise: instrumental (Gaussian and stationary)
 - Length of data set: **2 yr**
- **Note: data set provided now to enable the development of the necessary technique(s), but results expected in June 2007**



Data sets and resources

- Data sets are generated using LISA Simulator and Synthetic LISA (are they available to everyone):
 - Challenge data set: TDI variables (X,Y,Z)
- Signals parameters are drawn randomly from a range of values by means of simple codes and then used to generate h_+ and h_x (input for the simulators)
 - Task force decides on parameter range in order to ensure the generation of signals with given properties
 - Parameter range and distribution (with relevant codes) are public
 - The source codes to generate h_+ and h_x are public
- One member of the taskforce who **does not** take part in the challenges generates the actual parameters used in the challenge, h_+ and h_x and ultimately the data sets
 - For Challenge-1: John Baker
 - Codes are tagged, parameters stored under version control, etc.
- **Data, codes, documentation and relevant information are available at <http://astrogravs.nasa.gov/docs/mldc/>**



Reporting results

- Results:
 - Source(s) in the data stream
 - Parameters that characterise the source(s) in the form of posterior probability density functions, or confidence intervals, or any statistical quantity that is relevant for the adopted analysis strategy
- A technical note describing
 - Analysis method
 - Pipeline, software implementation, platform(s) on which the analysis was carried out and quantification of run times
- Participating groups are strongly encouraged (if at all possible) to tag the codes used for "production analysis" and maintain them under version control
- It is at the participant discretion to make the software available; although the TaskForce encourages this approach, this is not a requirement to participate to the MLDCs.



C1: schedule and organisation

- Full information are/will be available at:
<http://astrogravs.nasa.gov/docs/mldc/>
- Release date: 30th June 2006
 - Please [subscribe](#) to the challenge
 - Contact through mailing list and website
 - Please post [short note about plans](#)
- Extended teleconference around the beginning of October to track progress
- Due date: 1st December 2006
 - Submit results at <http://astrogravs.nasa.gov/docs/mldc/>
 - Extended teleconference ~ 5-8 December 2006
- GWDAW-11 at Albert-Einstein-Institut (Germany) 18th - 21st December, 2006
 - Face-to-face meeting
 - Dedicated presentation session at the workshop



Dissemination of results

- Results of MLDCs:
 - Technical note by each participating group
 - Technical note on summary of results issued jointly by TaskForce & participating groups
 - Proceedings of GWDAW by each participating group
 - Joint paper in international peer reviewed journal by TaskForce & participating groups [most likely for Challenge-2 and beyond]
- Analysis methods:
 - We expect papers about methods
 - It is important to be able to compare different analysis approaches and/or different implementations of the same analysis strategy
 - Training data sets
 - One can generate different realizations of the MLDC data sets

Conclusions

- Please take part in the MDLCs !
- Stay tuned for next talk by Michele Vallisneri

<http://astrogravs.nasa.gov/docs/mldc/>



LISA International Science Team (LIST)

ESA membership

- K Danzmann (co-chair)
- P Binetruy
- M Cerdonio
- M Cruise
- J Hough
- P Jetzer
- Y Mellier
- B Schutz
- T Sumner
- J-Y Vinet
- S Vitale

LIST website <http://www.srl.caltech.edu/lisa/>